Infrared Aero-Thermal Nonuniform Correction via Deep Multiscale Residual Network

Yi Chang, Student Member, IEEE, Luxin Yan, Member, IEEE, Li Liu, Houzhang Fang, and Sheng Zhong

Abstract—In the infrared focal plane arrays (IFPA) imaging systems, the temperature-dependent nonuniformity effects severely degrade the image quality. In this letter, we propose a very deep convolutional neural network for unified infrared aero-thermal nonuniform correction. Our network is built with the multiscale and residual training. The multiscale sub-networks utilize the multiscale property in the images, and the long-term residual learning contributes to the information propagation. Compared with the previous methods, the proposed method is more robust to various nonuniform artifacts, and more efficient at processing time. Experimental results validate the superiority of our method for infrared nonuniform correction.

Index Terms—Nonuniform correction, convolutional neural network, and infrared image.

I. INTRODUCTION

F OR the infrared imaging systems equipped on the high-speed aircraft, due to the temperatures fluctuations, the resulting images mainly contain two kinds of fixed pattern noise (FPN): a smooth nonuniform bias field which looks like bright and large spot, and line pattern nonuniform stripe noise, as shown in Fig. 1(a). This aero-thermal nonuniform effect severely influences the image quality for subsequent application. Therefore, it is necessary for us to remove these artifacts before the succeeding image interpretation processes are performed. In this letter, we mathematically formulate the degradation as follow:

\[ Y = X + B + E + N, \]  

where \( Y \in \mathbb{R}^{R \times C} \) is the observed image, \( R \) and \( C \) stand for the number of the rows and columns respectively, \( X \) is the clear image [Fig. 1(c)], \( B \) is the bias field [Fig. 1(c)], \( E \) is the stripe noise [Fig. 1(b)] (\( B \) and \( E \) are the nonuniform FPNs), and \( N \) is the random noise [Fig. 1(d)]. The goal of our work is to obtain the clear image \( X \) from the degraded image \( Y \).

The bias field is highly related to optical window temperature. The bias caused by thermal radiation looks like a bright and smooth spot, which is very similar to 2D Gaussian distribution. For the bias field removal, we mainly introduce the optimization based restoration methods [1]–[5]. Based on the correction model [6]. Cao et al. [2] locally fitted the derivatives of correction model to the gradient components with a subsequent bilateral filter for refinement. In [1], Zheng et al. firstly proposed an image decomposition based dual \( L_1 \) sparsity representation constraints for both the image and bias field component, respectively. Most of the optimization based methods followed this framework. Later, Liu et al. [5] introduced the sparser \( L_p \) gradient regularization for the images with better performance.

The stripe noise is mainly caused by the nonuniform response of adjacent detectors. The stripe noise has significantly directional characteristic due to its imaging mechanism. For the FPN stripe removal, there exist various kinds of methods: reference based approaches [7], filtering based methods [8], [9], the scene based optimization methods [10]–[12], learning based method [13], [14]. In [7], the authors proposed an algebraic-based algorithm which assumed that each IFPA detector output obeys an approximate linear irradiance voltage model. The authors [9] took advantage of the directional characteristic of the FPN and proposed an 1-Dimensional guided filter for stripe noise removal in infrared images. The optimization based sparsity methods have been popular in recent years. Vera et al. [10] proposed an isotropic total variation approach making use of an alternating minimization strategy for FPN removal. Further, the \( L_p \) gradient based iterative adaptive nonuniform correction method has been proposed [11], which assumed the gradient of the image is much more sparser. Recently, Kuang et al. [13] presented a three layer deep convolutional neural network (CNN) for single infrared image stripe noise removal.

Although numerous FPN correction methods have been
proposed in the past decades, all of them are designed for one specific task only, such as the nonuniform bias field \cite{1}–\cite{5} or nonuniform stripe noise \cite{7}, \cite{9}–\cite{14}. Moreover, most of the previous works utilized the hand-crafted features, such as the gradient or the dictionary coefficients. These hand-crafted priors only explore the locally shallow feature information of the images while ignoring the global high-level feature of the images. To overcome these limitations, we propose a deep multiscale residual network for infrared nonuniform correction. We introduce a very deep CNN model for extracting the high-level contextual information of the images, in which the discriminative features are benefit to distinguish each component in \cite{1} from each other. Our model does not rely on any pre-defined statistical assumption of the FPN or the image, which makes it very robust for arbitrary degradations. In addition, we utilize the multiscale information in the images via the CNN for better representation of both the large scale edge and fine texture in the images. Overall, the contributions of this work are as follows:

- This letter proposes a very deep CNN model for unified infrared nonuniform bias field and stripe correction. Compared with the previous methods, the deep features are more representative for both the IFPA image and artifacts.
- We introduce the long-short term residual learning strategy, which significantly reduces the difficulty of training. In addition, the multi-scale network could utilize the multi-scale property in the images, meanwhile, it can obviously reduce the computational load and memory.
- The proposed method outperforms the state-of-the-art methods by a large margin in terms of the speed, qualitative and quantitative assessments. Moreover, our method is very robust to the random noise in IFPA images.

II. DEEP MULTISCALE RESIDUAL NETWORK

Here, we will give our unified deep multiscale residual network (DMRN) for FPN removal. Although there are several deep learning based methods for infrared stripe noise removal, the depth of them is too shallow, such as 3 layers in \cite{13} and 10 layers in \cite{14}. On the contrary, we take the multiscale information of the image into consideration, and enlarge the receptive field via the multiscale feature extraction module.

A. MULTISCALE FEATURE EXTRACTION

It is well-known that the images contain different scales of information, such as the large scale edges and the fine textures. For one hand, the fine textures can be well represented by the shallow features with a relative small receptive field. And the large scale edges can be globally captured by the deep features with a relative larger receptive field. This motivates us to build a very deep network to globally capture both the local fine texture and global large edges. On the other hand, we argue that the multiscale structural information, which is benefit for image representation, need to be explicitly modeled by the different scales of feature maps. This motivates us to construct a multiscale pyramid network to explicitly depict the different scale structures.

The deep pyramid strategy has been extensively used in image segmentation \cite{15}, image restoration \cite{16}, due to its powerful representation ability. In this letter, we introduce a deep multiscale residual network for infrared nonuniform FPN removal, as shown in Fig. 2. We firstly downsample the image gradually via the convolutional layer with the stride 2. In this work, we extract 4 scale features from the original size 256*256 to 32*32. This part can be regarded as an encoder to extract the compact and multiscale representations. Then, we gradually upsample the image via the deconvolution layer with the scale 2. This part can be regarded as a decoder to reconstruct the signal. And the skip connection is introduced, where all the features with the same size are concatenated together, to improve the information flow among different layer. We also introduce the residual blocks \cite{17}, which create short paths among neighborhood features, to alleviate gradient vanishing and train a very deep model.

To leverage the multiscale information to guide the nonuniform correction, a connectivity fusion layer is further introduced, where the multiscale features from decoders are fused together. The intuitions behind this are two folds. For one hand, fusing the low-resolution features with the high-resolution features actually creates an information flow for better information propagation between low and high levels. On the other hand, this could better compensate low-level fine details to high level large scale features with joint and powerful representation for the images.
B. Reconstruction

In the second stage, the extracted multiscale features are the input of the reconstruction module which is used to obtain the clear image. Here, we use two convolutional blocks to achieve this goal. Note that, we do not learn the mapping of the clear image directly. We introduce the residual learning strategy by adding a skip connection between the input and output. Thus, the network actually learns the whole error ($B + E + N$), it guarantee that sparser gradient of the residual errors are easier to propagate. To reconstruct the image, we introduce the $L_2$ based loss function:

$$J = \frac{1}{2} || F(Y; W) - (B + E + N)||^2,$$

where $W$ is the mapping parameters to be learned. We introduce both the resblock (short term residual) and residual learning (long term residual) to increase the representation ability of the network. The residual learning greatly improves the depth of the network and avoids gradient vanishing issue.

C. Training Details

The MatConnet toolbox [18] is employed to train the model. The training code and IR datasets of our DMRN has be released at the homepage of the author[1]. We initialize the convolutional filters with Xavier method [17]. The learning rate starts from 0.0005 and is divided by 2 after each 20 epoch. The momentum and decay are fixed as 0.9 and 0, respectively. Adam solver [19] is used as optimization algorithm with a mini-batch size of 48. We train the model with 100 epoches. We randomly choose 10000 samples from the Place2 dataset with size 256*256 for pre-training. Then, we fine-tune the pretrained models on the collected 1500 mid-wave IR images.

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental Setting

For the bias field removal, we compare with the gradient components based filtering method (GCF) [2] and TV based decomposition methods (TVD) [3]. For the stripe noise removal, we compare with the LRSID [12] and DLS-NUC [14]. We use the codes provided by the authors, and fine-tune the hyper-parameters by default to achieve the best performance. The PSNR and SSIM are employed for the quantitative index. The visual correction and cross profile are used as the qualitative assessment. Due to the page limitation, more results and analysis are placed in the supplementary.

B. Comparison With STOA

1) Bias Field Correction: We compare DMRN with state-of-the-art methods for nonuniform bias field correction in the
IR images. In Fig. 3, the GCF and TVD always introduce unexpected artifacts (marked by the red ellipse). The results of proposed method are almost the same as the original image. The DMRN+ in Table I denotes the fine-tuned DMRN model. Under different conditions, the proposed DMRN consistently outperforms the state-of-the-art bias field correction methods.

2) Stripe Noise Removal: We compare DMRN with state-of-the-art methods for nonuniform stripe noise correction in the IR images. Here, we test them on different stripe noise levels and larger image size 480*480, as shown in Fig. 4. We can observe there exist obvious residual stripe noises in the correction results of LRSID and DLS-NUS as marked by the red ellipse, while in Fig. 4(e) the corrections results by DMRN are more visually pleasing. The quantitative comparison results are shown in Table II. The DMRN is robust to different noise level and image size, and consistently outperforms the state-of-the-art IR stripe correction methods.

C. Discussion

1) Cross Profile Analysis: In this subsection, we analyze the cross profile of the correction result, as shown in Fig. 5. We take the stripe noise correction [first row in Fig. 4] as an example. From the zoom part in Fig. 5(b), we can clearly observe that the correction result of DMRN (red curve) is much more closer to the original ground truth (black curve). Moreover, compared with the green curve of the striped image, we can infer from the smoothed red curve that the nonuniform stripe noise has been satisfactorily removed.

2) Joint Nonuniform Correction: We further demonstrate that our model is not only suitable for one specific nonuniform artifacts in IR, but also works well for arbitrary mixed nonuniform artifacts, due to the universal approximation theory of the deep neural network [20]. We test the DMRN on the mixed stripe and bias field, which are commonly seen nonuniform artifacts in IR image, along with the Gaussian random noise, as shown in Fig. 6. As far as we know, there is no way to uniformly correct all of these artifacts. The DMRN [Fig. 6(c)] can well remove the artifacts and preserve the edge structure.

3) Effectiveness of the Fine-tuning: To alleviate the lack of IR image, we firstly train our model on the RGB image. Then, we fine-tune our model on the collected IR images. We show this transfer learning strategy is very effective for nonuniform correction. In Fig. 8 the training loss and PSNR value of each
epoch are shown before and after the fine-tuning. It can be seen that the training loss drops suddenly at the epoch we fine-tune the model, and the loss of the fine-tuned model is significantly lower than the no fine-tuned model. On the contrary, the PSNR value increases after we fine-tune the model.

4) Benefit For Recognition: To validate the effectiveness of the proposed method for subsequent application, we employ Google Vision API on the images before and after correction to perform the scene recognition. Here, we choose a natural image as an example, since the API is mainly trained on natural images. As shown in Fig. 7, the bias field on the image puzzles the API with wrong labels such as light and sky. The recognition result of the correction by DMRN is almost the same as that of the original, such as the roof and house.

IV. CONCLUSION
In this letter, we propose a deep multiscale residual network for infrared image nonuniform correction. We utilize the deep feature which is more representative and robust to various nonuniform artifacts in IR image. The multiscale information is used in our network to better represent the IR images. The residual learning and fine-tuning strategy are introduced for better training the network. Experimental results show that the proposed DMRN is superior to competing deep and non-deep methods by a large margin.

REFERENCES

https://cloud.google.com/vision/
Supplementary

I. INTRODUCTION

II. FLOW CHART OF THE PROPOSED DMRN

The flow chart of the proposed method is shown in Fig. 1. In the training stage, first, we prepare the clean/degraded image pairs for training. Next, initialize the parameter $W$. Then, we pre-train our model on the natural image. Finally, we fine-tune our model on the infrared image and output the parameter $W$ for testing. In the testing stage, we use the trained parameter $W$ for DMRN inference.

III. MORE VISUAL RESULTS

Due to the page limitation, we just show several experimental results in our manuscript. However, as we have stated in our work, we have performed the comparison experiments on the collected 20 IR images. The average PSNR and SSIM values under different noise levels are reported in Table 1 and Table 2 in the manuscript. Here, we show more visual results of the DMRN on more natural images, as shown in Fig. 5.

IV. REAL EXPERIMENTAL RESULTS

Since the real bias field IR images are hard to find, we mainly show the mixed stripe noise and random noise correction results, as shown in Fig. 3. As you can see, the DMRN could remove the real stripe noise, at the same time well preserve the image structure. Compared with the DLS-NUS, the DMRN could well suppress the random noise with more visual pleasure appearance.

![Flow chart of the proposed method.](image-url)
V. Robustness to Other Images

The proposed model can be applied to any images with different imaging bands. On one hand, the output of the model is the residual error, such as the stripe noise and bias field not the image. That is to say, the trained model mainly focuses on extract the features of the errors. On the other hand, the result images with different imaging systems all contain similar edge and structures with difference in details. That is the main reason why we first train our model on natural image and fine-tune on infrared images. The learned features on the large natural image dataset can be well transferred to the infrared image. In Table I (in the main manuscript), we have shown the quantitative results of both DMRN and DMRN+ (fine-tuned model on DMRN). You can see that our model can work well for both the natural image and IR image. For example, the stripe noise is ubiquitous in multi-detector imaging system such as the hyperspectral image. We have shown our DRMR works well for the stripe noise removal in HSI, as shown in Fig. 2.

VI. Robustness to Different Image Size

We have tested the DMRN on different image size, as shown in Fig. 4. We degrade the image with joint nonuniform in the second column. The corresponding correction results are shown in the third column. We can observe that the correction results are consistently visual pleasure. In fact, we have already test our algorithm on large image in Fig. 4(in main manuscript). The image size is 480*480.

VII. Effectiveness of the Residual Net

This residual net figures out to learn the sparse residual image, not the image itself, since the sparser gradient of the residual image is easier to be propagated. The residual learning greatly improves the depth of the network and avoids gradient vanishing issue. In our work, we introduce both the resblock (short term residual) and residual learning (long term residual) to increase the representation ability of the network. Here, we compare the quantitative results (stripe noise level $S \subset \{-50, 50\}$) between DMRN model with/without residual strategy, as shown in Table II. We can observe that the proposed DMRN obtains better results than the DMRN without the residual net. This experiment validates the effectiveness of the residual net.

VIII. Running Time

In Table I, we list the running time (inference) comparison with the competing methods. As you can see, the CNN-based methods are obviously faster than the conventional methods no matter the CPU or GPU version. The running time of DMRN is slower than that of DLS-NUS, since our network is much more deeper and larger.
Table I

**Quantitative comparison between with and without residual net.**

<table>
<thead>
<tr>
<th>Method</th>
<th>Index</th>
<th>Img1</th>
<th>Img2</th>
<th>Img3</th>
<th>Img4</th>
<th>Img5</th>
<th>Img6</th>
<th>Img7</th>
<th>Img8</th>
<th>Img9</th>
<th>Img10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without</td>
<td>PSNR</td>
<td>33.27</td>
<td>33.86</td>
<td>34.09</td>
<td>33.58</td>
<td>30.68</td>
<td>31.33</td>
<td>30.99</td>
<td>34.90</td>
<td>34.09</td>
<td>32.86</td>
</tr>
<tr>
<td>Residual</td>
<td>SSIM</td>
<td>0.9748</td>
<td>0.9649</td>
<td>0.9727</td>
<td>0.9726</td>
<td>0.9528</td>
<td>0.9509</td>
<td>0.9478</td>
<td>0.9776</td>
<td>0.9702</td>
<td>0.9616</td>
</tr>
<tr>
<td>With</td>
<td>PSNR</td>
<td>34.64</td>
<td>35.35</td>
<td>36.05</td>
<td>35.13</td>
<td>32.13</td>
<td>33.30</td>
<td>32.74</td>
<td>36.88</td>
<td>35.29</td>
<td>33.79</td>
</tr>
<tr>
<td>Residual</td>
<td>SSIM</td>
<td>0.9836</td>
<td>0.9767</td>
<td>0.9832</td>
<td>0.9831</td>
<td>0.9697</td>
<td>0.9701</td>
<td>0.9648</td>
<td>0.9854</td>
<td>0.9786</td>
<td>0.9716</td>
</tr>
</tbody>
</table>

Table II

**Running time (second) comparison under different image size.**

<table>
<thead>
<tr>
<th>Size</th>
<th>GCF</th>
<th>TVD</th>
<th>LRSID</th>
<th>DLS-NUS(CPU)</th>
<th>DLS-NUS(GPU)</th>
<th>DMRN(CPU)</th>
<th>DMRN(GPU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>256*256</td>
<td>3.24</td>
<td>2.90</td>
<td>4.15</td>
<td>0.09</td>
<td>0.0046</td>
<td>0.22</td>
<td>0.0125</td>
</tr>
<tr>
<td>480*480</td>
<td>22.22</td>
<td>10.66</td>
<td>14.83</td>
<td>0.36</td>
<td>0.0088</td>
<td>0.71</td>
<td>0.0133</td>
</tr>
<tr>
<td>Light Error</td>
<td>Degrade</td>
<td>DMRN</td>
<td>Error</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>---------</td>
<td>------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(PSNR, SSIM)</td>
<td>(23.06, 0.5843)</td>
<td>(33.41, 0.9270)</td>
<td>(36.32, 0.8870)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heavy Error</td>
<td>(PSNR, SSIM)</td>
<td>(11.93, 0.3803)</td>
<td>(28.43, 0.8546)</td>
<td>(31.33, 0.9278)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Light Bias</td>
<td>(PSNR, SSIM)</td>
<td>(25.10, 0.9867)</td>
<td>(77.37, 1.000)</td>
<td>(64.61, 0.9995)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heavy Bias</td>
<td>(PSNR, SSIM)</td>
<td>(13.15, 0.7697)</td>
<td>(34.35, 0.9969)</td>
<td>(38.58, 0.9896)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Light Stripe</td>
<td>(PSNR, SSIM)</td>
<td>(32.89, 0.8908)</td>
<td>(44.87, 0.9973)</td>
<td>(48.13, 0.9749)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heavy Stripe</td>
<td>(PSNR, SSIM)</td>
<td>(19.35, 0.3799)</td>
<td>(38.88, 0.9912)</td>
<td>(41.88, 0.9900)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. More visually experimental results of DMRN on natural images. We show the joint and individual correction results under different noise level.
Figure 3. Real IR mixed noise correction. First column is the degrade images. Second column is the results of DLS-NUS. Third column is the results of DMRN.
Figure 4. Effectiveness for hyperspectral image. (a) Original HSI, (b) Simulated stripe image, (c) DMRN.
<table>
<thead>
<tr>
<th>Image Size</th>
<th>Original</th>
<th>Degrade</th>
<th>DMRN</th>
</tr>
</thead>
<tbody>
<tr>
<td>128×128</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td>256×256</td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>512×512</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
<tr>
<td>1024×1024</td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
</tbody>
</table>

Figure 5. Robustness for different size images. We test DMRN on image size from 128×128, 256×256, 512×512, 1024×1024.